Latent Capacity Region: A Case Study on Symmetric Broadcast With Common Messages

Chao Tian, *Member, IEEE*

***Abstract*—We consider the problem of broadcast with common messages, and focus on the case that the common message rate $R_{\mathcal{A}}$, i.e., the rate of the message intended for all the receivers in the set $\mathcal{A}$, is the same for all the set $\mathcal{A}$ of the same cardinality. Instead of attempting to characterize the capacity region of general broadcast channels, we only consider the structure of the capacity region that any broadcast channel should bear. The concept of latent capacity region is useful in capturing these underlying constraints, and we provide a complete characterization of the latent capacity region for the symmetric broadcast problem. The converse proof of this tight characterization relies on a deterministic broadcast channel model. The achievability proof generalizes the familiar rate transfer argument to include more involved erasure correction coding among messages, thus revealing an inherent connection between broadcast with common message and erasure correction codes.**

***Index Terms*—Broadcast channel, common message, individual message.**

I. Introduction

One central theme in multi-user information theory (IT) is the pursuit of single-letter[1] characterizations of the capacity regions for channel coding problems, or the achievable rate regions (possibly under certain distortion constraints) for source coding problems. However, some useful properties of these regions can be identified, e.g., convexity, even when a single-letter characterization is not available. An immediate question to ask is whether there exist other properties of the capacity region that do not rely on a single letter characterization.

The following question is of interest in this regard: in a particular multi-user IT problem, can the achievability of a rate vector $(R_1^*, R_2^*, \ldots, R_N^*)$ imply the achievability of any rate vector in some region[2] $\mathcal{R}(R_1^*, R_2^*, \ldots, R_N^*)$, regardless of the exact probabilistic channel model? We show that indeed this is true for the symmetric broadcast problem, and this region can be rather non-trivial. We denote the largest of such regions $\mathcal{R}(R_1^*, R_2^*, \ldots, R_N^*)$ as $\mathcal{C}(R_1^*, R_2^*, \ldots, R_N^*)$ in a channel coding problem, and call it the *latent capacity region* implied by $(R_1^*, R_2^*, \ldots, R_N^*)$; the *latent achievable rate region* can be similarly defined, possibly under certain distortion constraints, for a source coding problem though it is not our main focus.

Chao Tian is with AT&T Labs-Research, Florham Park, NJ 07932, USA (email: tian@research.att.com).

[1]The emphasis on single letter is largely because such kind of characterization is usually computable.

[2]Apparently the region defined by $R_i \leq R_i^*$ is implied in a channel coding problem, but this trivial case is not interesting. Note here we do not take the subscript of rate R_i^* to have any specific meaning associated with the user indices, but merely as an integer label to enumerate the rates in question.

For broadcast and multiple access channels, a precise problem formulation was given in a recent work by Grokop and Tse [1], called *multicast region*, which provides a framework to answer the above question. Complete solutions were found in [1] for broadcast and multiple access channels with **two** and **three** users, but the problem remains open for more than three users. We believe this problem formulation reveals a more general concept not limited to only these two channels, and thus rename it as the latent capacity (or latent achievable rate) region problem to make explicit this generality. Our perspective is different from [1] in that we wish to highlight the importance of the latent capacity region concept in its "maximum implication" meaning, and thus we shall define the region in an alternative (but equivalent) manner to emphasize this perspective; our interest in this problem is partially due to an observation made during an earlier work [2], as we shall discuss shortly.

One may wonder how a single achievable rate vector $(R_1^*, R_2^*, \ldots, R_N^*)$ can imply the achievability of a certain region. In some cases, it is perhaps best explained by the familiar rate transfer argument, that the rate to transmit common messages can be used to transmit individual messages instead, and vice versa. For example, for a two user broadcast channel, if a common message rate $R_{\{1,2\}}^*$, and individual message rates $R_{\{1\}}^*$ and $R_{\{2\}}^*$ are achievable, respectively, then it is not difficult to see that the region of $(R_{\{1,2\}}, R_{\{1\}}, R_{\{2\}})$ given below is achievable by transferring between common and individual rates (see also [1])

$$\begin{aligned} R_{\{1,2\}} + R_{\{1\}} &\leq R_{\{1,2\}}^* + R_{\{1\}}^* \\ R_{\{1,2\}} + R_{\{2\}} &\leq R_{\{1,2\}}^* + R_{\{2\}}^* \\ R_{\{1,2\}} + R_{\{1\}} + R_{\{2\}} &\leq R_{\{1,2\}}^* + R_{\{1\}}^* + R_{\{2\}}^*. \end{aligned}$$

However, for more than two users, such a naive rate transfer argument is not sufficient, and additional processing is needed, as observed in [1] for the three user case. In fact, this was exactly the perspective taken in [1], where the goal is to exhaust all such rate transfer operations. The perspective taken in [1] and that taken here are complementary to each other, and one may suit certain problems better than the other. Because of this relation, it is not surprising that the achievability proof of our result also relies on a generalized version of rate transfer operations. We shall show that when more users are involved, such generalized rate transfer operation requires strategic application of erasure correction codes, which reveals an inherent connection between erasure correction codes and broadcast with common messages. More specifically, in this work, we shall largely stay in the framework of [1], and provide a complete solution to the K-user broadcast channel

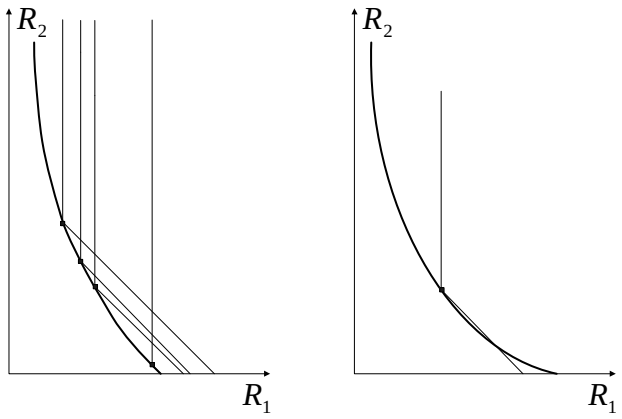

Fig. 1. The bold curve gives the rate region: while the left one is possible, the right one is impossible for the successive refinement source coding problem. The thin lines give the latent capacity region associated with each small black dot.

latent capacity region problem under an additional symmetry constraint, whereas only cases with two and three users were solved in [1] without such a constraint.

The characterization of latent capacity/rate region is important in multi-user IT for two reasons. First, it may facilitate finding a single-letter characterization or an approximate characterization. For example, a rate-distortion region characterization for the problem of multi-stage successive refinement with degraded decoder side information was given in the form of bounds on sum-rates [2] as

$$\sum_{i=1}^{m} R_i \geq \sum_{i=1}^{m} I(X; W_m | W_1, W_2, ..., W_{m-1}, Y_m), \quad 1 \leq m \leq N.$$

On the other hand, it seems impossible to establish directly the converse for a characterization in the form of bounds on each individual incremental rate [3], despite the fact that the two characterizations are equivalent [2]. This is not a coincidence, and it is not difficult to show that the latent rate region for this problem has exactly the following form, assuming non-negativity of the rates

$$\sum_{i=1}^{m} R_i \geq \sum_{i=1}^{m} R_i^*, \quad 1 \leq m \leq N. \tag{1}$$

Intuitively, when a rate vector $(R_1^*, R_2^*, \ldots, R_N^*)$ is achievable, its latent capacity/rate region gives the largest achievable region thus implied, i.e., maximally utilizes it, which may help simplify the representation of the region when taking the union over auxiliary random variables. Similarly, when an approximate characterization is needed, a good inner bound may be found by choosing one or several good (auxiliary coding) distributions in an information theoretic coding scheme which lead to one or several rate vectors, and then taking the convex hull of their latent rate regions. One simple example is in [4], where an approximate characterization for the side-information scalable source coding problem was given for general sources under the squared error distortion measure, and the inner bound approximation is exactly the latent capacity region implied by a single rate pair.

The second reason making this concept important is even if it does not lead to a single-letter characterization or an approximate characterization, it can still provide insights into the problem. One such example is that the capacity region can always be written as the (possibly uncountable) union of latent capacity regions, which places certain constraints on the geometry of the achievable region. For the above example of successive refinement source coding, we show in Fig. 1 a possible rate region on the left, and an impossible rate region on the right. The one on the right is impossible because the black dot is in the achievable region, thus the latent capacity region implied by it (given by the thin line) must be also in the region, which is not satisfied by the region depicted on the right. This important observation was also discussed in [1] (see Corollary 4.3), and we do not elaborate it further. Nevertheless, it is rather clear that the latent capacity region indeed provides fundamental and useful property of the rate region, in addition to the well-known convexity.

II. Problem Definition and Preliminaries

We first define the symmetric broadcast problem, and then introduce the notion of latent capacity region in this context.

In a general K-user broadcast channel, the conditional probability distribution is given as

$$p(y_1[1,2,\ldots], y_2[1,2,\ldots], ..., y_K[1,2,\ldots] \big| x[1,2,\ldots]) \tag{2}$$

where the index in the bracket $[1,2,...]$ is used to denote time; the random variables have alphabets $\mathscr{X}, \mathscr{Y}_1, \ldots, \mathscr{Y}_2$, and the receivers are indexed as $1, 2, \ldots, K$. The alphabets can be discrete or continuous, and the channel can be memoryless or otherwise; for our purpose, it is perhaps beneficial, though not necessary, to limit the attention to cases where the channel transition process is (block) stationary and ergodic. We use script letters to denote sets, and particularly, $\mathcal{A}$ and $\mathcal{B}$ are reserved for subsets of $\mathcal{I}_K = \{1, 2, \ldots, K\}$, i.e.,

$$\mathcal{A}, \mathcal{B} \subseteq \{1, 2, \ldots, K\}. \tag{3}$$

$|\mathcal{A}|$ is used to denote the cardinality of set $\mathcal{A}$. A length-n vector $X[1,2,...,n]$ is sometimes written as X^n; for a K dimensional vector $(R_1, R_2, ..., R_K)$, we sometimes write it simply as $\boldsymbol{R}$.

Let $\{W_{\mathcal{A}}, \mathcal{A} \subseteq \mathcal{I}_K\}$ be 2^K mutually independent and uniformly distributed messages, where $W_{\mathcal{A}}$ is the message intended for all the receivers in the set $\mathcal{A}$; for notational convenience, we include $W_\emptyset$ but will assume it to be a constant. For each $k = 1, 2, \ldots, K$, define the set of random variables

$$\mathcal{W}_k = \{W_{\mathcal{A}}, \mathcal{A} : k \in \mathcal{A}\}. \tag{4}$$

Thus $\mathcal{W}_k$ is the collection of messages that the k-th receiver should decode. We also define the following set of random variables

$$\overline{\mathcal{W}}_k = \{W_{\mathcal{A}}, |\mathcal{A}| \geq k\}. \tag{5}$$

More specifically for $K=3$, we have

$$\begin{aligned}\mathcal{W}_1 &= \{W_1, W_{12}, W_{13}, W_{123}\}\\ \mathcal{W}_2 &= \{W_2, W_{12}, W_{23}, W_{123}\}\\ \mathcal{W}_3 &= \{W_3, W_{13}, W_{23}, W_{123}\}\\ \overline{\mathcal{W}}_1 &= \mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3\\ \overline{\mathcal{W}}_2 &= \{W_{12}, W_{13}, W_{23}, W_{123}\}\\ \overline{\mathcal{W}}_3 &= \{W_{123}\},\end{aligned} \tag{6}$$

where we have slightly abused the notation by writing, e.g., $W_{\{1\}}$ as W_1. The sets $\mathcal{X}_i^n$ and $\overline{\mathcal{X}}_i^n$ are defined similarly for length-n random vectors. In this work, we only consider the case that the rates of messages $W_{\mathcal{A}}$ are the same for all such messages where the set $\mathcal{A}$ has the same cardinality. More formally, the problem is defined as follows.

***Definition** 1:* An $(n, R_1, R_2, \ldots, R_K)$ symmetric broadcast code consists of an encoder

$$f: \prod_{\mathcal{A}\subseteq\mathcal{I}_K} \mathcal{I}_{2^{nR_{|\mathcal{A}|}}} \to \mathscr{X}^n, \tag{7}$$

where $R_\emptyset \triangleq 0$ and K decoders,

$$g_k: \mathscr{Y}_k^n \to \prod_{\mathcal{A}:k\in\mathcal{A}} \mathcal{I}_{2^{nR_{|\mathcal{A}|}}}, \tag{8}$$

resulting in the decoded messages at the k-th receiver $\{\hat{W}_{k,\mathcal{A}} : k\in\mathcal{A}\}$, and the decoding error probability of at least one message at one receiver

$$P_e^{(n)} = \Pr\left(\bigcup_{k=1}^{K}\bigcup_{\mathcal{A}:k\in\mathcal{A}} \{W_{\mathcal{A}} \neq \hat{W}_{k,\mathcal{A}}\}\right). \tag{9}$$

***Definition** 2:* A rate vector $\boldsymbol{R}$ is symmetrically achievable if there exists a sequence of $(n, \boldsymbol{R})$ codes with $P_e^{(n)} \to 0$. The closure of the set of symmetrically achievable rate vectors is called the symmetric broadcast capacity region, denoted as $\mathcal{C}_{p(y_1,y_2,\ldots,y_K|x)}$, or simply as $\mathcal{C}_p$.

Note that secrecy constraint is not considered in the definition. Next we define the latent capacity region for this problem.

***Definition** 3:* For a given rate vector $\boldsymbol{R}^*$, the collection of rate vectors $\mathcal{R}(\boldsymbol{R}^*)$ is called the latent capacity region for symmetric broadcast implied by $\boldsymbol{R}^*$, denoted as $\mathcal{C}(\boldsymbol{R}^*)$, if the following two conditions are satisfied (i) For any broadcast channel, $\boldsymbol{R}^* \in \mathcal{C}_p$ implies $\mathcal{R}(\boldsymbol{R}^*) \subseteq \mathcal{C}_p$; (ii) There exists a set of channels $\{p_x\}$, such that $\boldsymbol{R}^* \in \mathcal{C}_{p_x}$ and $\mathcal{R}(\boldsymbol{R}^*) \supseteq \bigcap_x \mathcal{C}_{p_x}$.

For the second condition, we essentially wish to find one particular channel such that $\mathcal{R}(\boldsymbol{R}^*) \supseteq \mathcal{C}_p$. However this does not quite serve the purpose since this channel might be difficult to realize, however it can always be approximated by a sequence of channels. The above definition is slightly different from the one in [1], which is

$$\mathcal{C}(\boldsymbol{R}^*) = \bigcap_{p:\boldsymbol{R}^*\in\mathcal{C}_p} \mathcal{C}_p. \tag{10}$$

It can be easily verified that they are equivalent. The problem we wish to solve is the characterization of $\mathcal{C}(\boldsymbol{R}^*)$. It is clear that the region $\mathcal{C}(\boldsymbol{R}^*)$ is uniquely defined for any $\boldsymbol{R}^*$, and thus the problem is meaningful.

Definition 3 makes clear the "maximal implication" meaning of the latent capacity region. In multi-user IT, usually a coding scheme is given by fixing some auxiliary random variables, and then showing a single rate vector is achievable with certain random codes; the task of maximizing the implication region of this single point is sometimes mingled with the conditions under which this single point is achievable. The concept of latent capacity region can be used to delineate them.

The following lemma is needed in the converse proof.

***Lemma** 1 (K-way submodularity):* Let $\{U_i, i = 1, 2, \ldots, N\}$ be a set of mutually independent random variables, and $\{V_i, i = 1, 2, \ldots, N\}$ be a set of random variables jointly distributed with it. Let $\mathcal{G}_i$, $i = 1, 2, \ldots, K$ be subsets of $\mathcal{I}_N$. Then

$$\sum_{k=1}^{K} H(V_i, i\in\mathcal{G}_k | U_i, i\in\mathcal{G}_k) \geq \sum_{k=1}^{K} H(V_i, i\in\widehat{\mathcal{G}}_k | U_i, i\in\widehat{\mathcal{G}}_k), \tag{11}$$

where

$$\widehat{\mathcal{G}}_k \triangleq \bigcup_{\{j_1,j_2,\ldots,j_k\}\subseteq\mathcal{I}_K} (\mathcal{G}_{j_1}\cap\mathcal{G}_{j_2}\cap\cdots\cap\mathcal{G}_{j_k}). \tag{12}$$

This lemma is a direct consequence of the sub-modularity of the conditional entropy function, when the random variables being conditioned on are independent (a proof is given in Appendix A), and the K-way submodularity property of any submodular function given in [5].

III. Main Result

Our main result is a complete characterization of the latent capacity region for the symmetric broadcast problem. To present this region, a few more quantities need to be defined first. Let us define the following up-exchange rate for $i<j$

$$\phi_{i,j} = \binom{K-i}{j-i}^{-1}\binom{j-1}{j-i}, \tag{13}$$

and the down-exchange rate for $i>j$

$$\phi_{i,j} = \binom{i}{i-j}^{-1}\binom{K-j}{i-j}, \tag{14}$$

and define $\phi_{i,i} = 1$. The up/down exchange rates $\phi_{i,j}$ essentially describe the ratio when converting certain type of messages into other types. For example when $K=3$, the common message W_{123} can be used to convey individual information to the three users, and vice versa, but the conversion of such rates is not always ratio one. It will become clear in the achievable proof how such conversion can be done in a most efficient manner.

Define $\mathcal{C}^*(\boldsymbol{R}^*)$ to be the set of rate vectors $\boldsymbol{R}$ satisfying the following conditions with some K^2 non-negative quantities

$r_{i,j}$, $(i,j) \in \mathcal{I}_K \times \mathcal{I}_K$,

$$R_i^* \geq \sum_{j=1}^{K} r_{i,j}, \quad i = 1, 2, \ldots, K, \tag{15}$$

$$0 \leq R_j \leq \sum_{i=1}^{K} \phi_{i,j} r_{i,j}, \quad j = 1, 2, \ldots, K. \tag{16}$$

Roughly speaking, the rate $r_{i,j}$ is that taken from level-i rate R_i^* but used to transmit level-j messages. We have the following theorem.

***Theorem** 1:* For any non-negative rate vectors $(R_1^*, R_2^*, \ldots, R_K^*)$, we have

$$\mathcal{C}(R_1^*, R_2^*, \ldots, R_K^*) = \mathcal{C}^*(R_1^*, R_2^*, \ldots, R_K^*). \tag{17}$$

Example: for $K = 2$, it is straightforward to see: $\phi_{1,2} = 1$, i.e., the same amount of individual message rate for each user can be used to transmit a common message; and $\phi_{2,1} = 1/2$, i.e., to split a common message into two equal parts, each to transmit a separate individual message for one user.

Example: for $K = 3$, it can be verified using Fourier-Motzkin elimination [7] that $\mathcal{C}^*(R_1^*, R_1^*, \ldots, R_K^*)$ is given by the non-negative rates satisfying

$$\begin{aligned}
3R_1 + 6R_2 + 2R_3 &\leq 3R_1^* + 6R_2^* + 2R_3^*, \\
2R_1 + 2R_2 + 1R_3 &\leq 2R_1^* + 2R_2^* + 1R_3^*, \\
1R_1 + 2R_2 + 1R_3 &\leq 1R_1^* + 2R_2^* + 1R_3^*, \\
3R_1 + 3R_2 + 1R_3 &\leq 3R_1^* + 3R_2^* + 1R_3^*.
\end{aligned} \tag{18}$$

A typical shape is given in Fig. 2 with $(R_1^*, R_2^*, R_3^*) = (1, 2, 2)$. The computation is tedious and thus omitted here. The same result can also be reduced from that given in [1] for the asymmetric case. It is clear that this region is non-trivial, and it is not at all clear a priori why these rate combinations should be considered.

In [1], the region is characterized by investigating the distinct universal encoding/decoding operations, which leads to the concept of extremal rays. Because the latent capacity region in question is a polytope, it can be characterized by its faces, edges, or vertices. The extremal rays are essentially the edges of this polytope. However this proof approach in [1] appears rather difficult to generalize for more than three users since the number of edges quickly becomes very large, and thus we introduce the parametric characterization (15) and (16) to avoid this difficulty.

Notice that the exchange rate is pairwise, suggesting in this symmetric setting there is no need to convert rates jointly, e.g., use W_{12} and W_3 to send the same message W_{123}. In the rest of the paper, we shall prove Theorem 1. The naive approach of finding the planes of the rate region and derive its upper and lower bounds is not appropriate for general K, particularly for the purpose of converse. Instead, we utilize the structure of the region $\mathcal{C}^*(\boldsymbol{R}^*)$ to give a proof.

IV. Proof of the Forward Part For Theorem 1

The proof of the forward part of Theorem 1, i.e., the fact that $\mathcal{C}^*(\boldsymbol{R}^*)$ satisfies the first condition in Definition 3 is relatively straightforward.

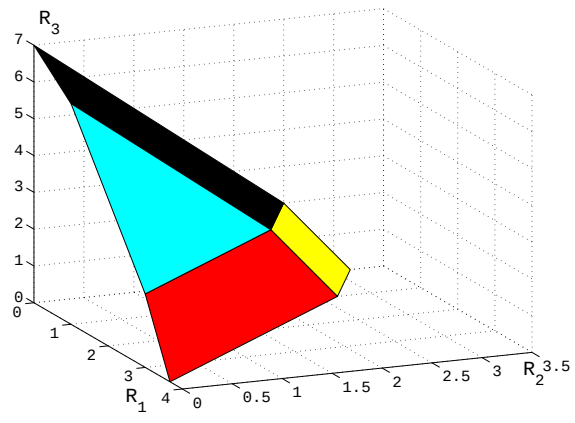

Fig. 2. The latent capacity region implied by rate vector $(1, 2, 2)$.

Proof of forward part for Theorem 1: Since $(\boldsymbol{R}^*)$ is achievable on any channel, there exists a sequence of codes with such rates with $P_e^{(n)} \to 0$, and we will use these codes to construct a set of codes to approach any rate vectors in $\mathcal{C}^*(\boldsymbol{R}^*)$. This is done by essentially relabeling and adding erasure correction codes on the messages.

Observe that the messages $\{W_{\mathcal{A}}, |\mathcal{A}| = i\}$ can also be used to transmit common messages to the subsets with cardinality smaller or larger than i. Moreover, we can use part of the rate R_i^*, denoted as $r_{i,j}$, for this purpose, to transmit some messages $\{W'_{\mathcal{A}}, |\mathcal{A}| = j\}$, thus increasing R_j. Such an operation will cause a conversion of rate $r_{i,j}$ for i-user subset messages into rate $\phi_{i,j} r_{i,j}$ for j-user subset messages, with an exchange rate $\phi_{i,j}$. The region $\mathcal{C}^*(\boldsymbol{R}^*)$ is precisely the result of allowing this kind of pairwise exchange on the rate vector $\boldsymbol{R}^*$. Thus we only need to show that the exchange rates $\phi_{i,j}$ given before Theorem 1 is indeed valid, then the existing sequence of channel codes can be used directly.

It is clear that we only need to consider the following problem: on a channel with $R_i = R$ and $R_k = 0$ for $k \neq i$, how do we transmit messages $\{W_{\mathcal{A}}, |\mathcal{A}| = j\}$, and how much rate R_j can be supported? We will only need to distinguish two cases $i < j$ or $i > j$, since it is clear $\phi_{i,i} = 1$.

We first consider the case $i < j$. For a subset $\mathcal{B}$ of $\mathcal{I}_K$ where $|\mathcal{B}| = j$, there are a total of $\binom{j}{i}$ subset of $\mathcal{B}$ with cardinality i; denote the collection of such subsets as $2^{\mathcal{B},i}$. For a particular user $k \in \mathcal{B}$, it can decode (with high probability) the messages $\{W_{\mathcal{A}} : k \in \mathcal{A} \subset \mathcal{B}\}$, i.e., $\binom{j-1}{i-1}$ such messages. To transmit the common message $W_{\mathcal{B}}$, if we can guarantee that when receiving any $\binom{j-1}{i-1}$ messages out of the $\binom{j}{i}$ messages in the set $2^{\mathcal{B},i}$, the message is decodable, then it is clear that indeed any receivers in the set $\mathcal{B}$ can decode the message $W_{\mathcal{B}}$. This is an erasure correction problem and a $\left(\binom{j}{i}, \binom{j-1}{i-1}\right)$ maximum distance separable (MDS) code can satisfy this requirement, which indeed exists when the codeword length is sufficiently large. Furthermore, since each subset $\mathcal{A}$ of cardinality i is a subset of $\binom{K-i}{j-i}$ sets of cardinality j, only $\binom{K-i}{j-i}^{-1}$ of the rate

$R_{\mathcal{A}}$ can be used for each MDS code. This yields

$$R_j = \binom{K-i}{j-i}^{-1}\binom{j-1}{i-1}R_i = \phi_{i,j}R_i. \tag{19}$$

Next consider the case $i > j$. Let $\mathcal{B}$ be a subset of $\mathcal{I}_K$ where $|\mathcal{B}| = i$. The common message $\mathcal{B}$ can be shared uniformly between its $\binom{i}{j}$ subsets of cardinality j, for transmitting their "individual" message. Since each subset $\mathcal{A}$ of cardinality j is a subset of distinct $\binom{K-j}{i-j}$ sets of cardinality i, it can take part in such sharing $\binom{K-j}{i-j}$ times. This yields

$$R_j = \binom{i}{j}^{-1}\binom{K-j}{i-j}R_i = \phi_{i,j}R_i. \tag{20}$$

Taking into account the existence of good MDS code, and the fact that $\mathcal{C}_p$ is a closed set, the proof is complete. ■

In [1], it was observed that in order to efficiently transfer rates, sometimes a modulo two addition is needed, similar to that seen in butterfly network of network coding [6]. The MDS codes we use in the above proof can be understood as a generalization of the modulo two addition, which itself is essentially a $(3, 2)$ MDS code. It is worth noting that other coding/processing may also be useful for converting rates, however, MDS codes are sufficient in solving the symmetric broadcast problem.

V. Proof of the Converse Part For Theorem 1

The converse proof of Theorem 1 requires more work. For simplicity we shall assume $2^{R^*_{\mathcal{A}}}$'s are all integers; if this is not the case, a sequence of channels need to be considered, and we shall return to this technical point after the proof.

We only need to provide one particular channel that $\boldsymbol{R}^* \in \mathcal{C}_p$ and $\mathcal{R}(\boldsymbol{R}^*) \supseteq \mathcal{C}_p$. The channel is the deterministic one considered in [1], extended to the K-user case; see Fig. 3 for the case $K = 3$. More precisely, let the channel input be the collection of $\{X_{\mathcal{A}}, \mathcal{A} \subseteq \mathcal{I}_K\}$. The alphabet of $X_{\mathcal{A}}$ where $|\mathcal{A}| = k$ is $\mathcal{I}_{2^{R^*_k}}$. The k-th channel output Y_k is given by

$$Y_k = \{X_A : k \in \mathcal{A}\}. \tag{21}$$

Denote this deterministic channel as p^*. In order to prove the converse part for Theorem 1, we need to establish $\mathcal{C}^*(\boldsymbol{R}^*) \supseteq \mathcal{C}_{p^*}$ for this channel.

For any $\boldsymbol{A} = A_1, A_2, \ldots, A_K$ where $A_i \geq 0$, define the following quantity

$$B_{\mathcal{C}^*}(\boldsymbol{A}) = \max_{\boldsymbol{R}\in\mathcal{C}^*(\boldsymbol{R}^*)} \sum_{k=1}^{K} A_k R_k, \tag{22}$$

and similarly

$$B_{\mathcal{C}}(\boldsymbol{A}) = \max_{\boldsymbol{R}\in\mathcal{C}_{p^*}} \sum_{k=1}^{K} A_k R_k. \tag{23}$$

It is clear that both $\mathcal{C}^*(\boldsymbol{R}^*)$ and $\mathcal{C}_{p^*}$ are convex regions, and thus if we can prove the following theorem, then the converse of Theorem 1 directly follows.

***Theorem** 2:* For any $\boldsymbol{A}$ where $A_i \geq 0$,

$$B_{\mathcal{C}^*}(\boldsymbol{A}) \geq B_{\mathcal{C}}(\boldsymbol{A}). \tag{24}$$

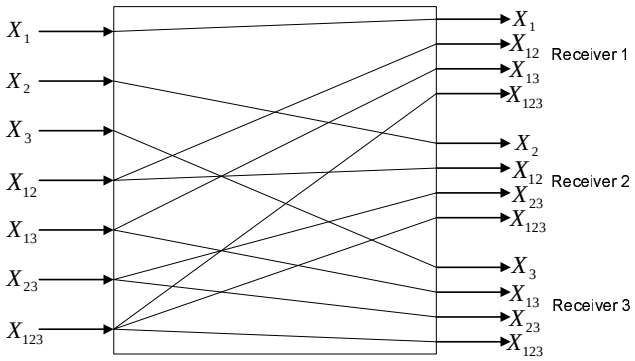

Fig. 3. The deterministic broadcast channel $K = 3$ from [1].

This is indeed our proof approach, however before giving the rather long proof for the general case, we first prove a few rate combinations for $K = 3$, which illustrates the basic techniques as well as facilitates better understanding. Though the proof of the case for $K = 3$ can also be found in [1], our proof given here is different and in fact more structured, which is geared toward the general case. After this example, a few necessary tools and intermediate results are provided, and finally we give the converse proof of Theorem 1.

A. Bounding Two Rate Combinations for $K = 3$

We give an outline of the proof for the first two inequalities in the example given after Theorem 1.

Proof:

$$\begin{aligned}
&3nR_1^* + 6nR_2^* + 2nR_3^* \\
&\geq \frac{2}{3}\sum_{i=1}^{3} H(\mathcal{X}_i^n) + \frac{1}{3}\sum_{i=1}^{3} H(\mathcal{X}_i^n|X_{123}^n, W_{123}) \\
&\overset{(a)}{\geq} \frac{2}{3}\sum_{i=1}^{3}[H(\mathcal{X}_i^n|\mathcal{W}_i) + H(\mathcal{W}_i)] \\
&\qquad + \frac{1}{3}\sum_{i=1}^{3} H(\mathcal{X}_i^n|X_{123}^n, W_{123}) - n\delta \\
&\overset{(b)}{=} 2nR_1 + 4nR_2 + 2nR_3 + \frac{2}{3}\sum_{i=1}^{3} H(\mathcal{X}_i^n|\mathcal{W}_i) \\
&\qquad + \frac{1}{3}\sum_{i=1}^{3} H(\mathcal{X}_i^n|W_{123}) - H(X_{123}^n|W_{123}) - n\delta \\
&\overset{(c)}{\geq} 2nR_1 + 4nR_2 + 2nR_3 + \frac{2}{3}H(X_{123}^n|W_{123}) \\
&\qquad + \frac{1}{3}\sum_{i=1}^{3}[H(\mathcal{W}_i|W_{123}) + H(\mathcal{X}_i^n|\mathcal{W}_i)] \\
&\qquad - H(X_{123}^n|W_{123}) - n\delta' \\
&= 3nR_1 + 6nR_2 + 2nR_3 - n\delta' \\
&\qquad + \left[\frac{1}{3}\sum_{i=1}^{3} H(\mathcal{X}_i^n|\mathcal{W}_i) - \frac{1}{3}H(X_{123}^n|W_{123})\right] \\
&\overset{(d)}{\geq} 3nR_1 + 6nR_2 + 2nR_3 - n\delta',
\end{aligned} \tag{25}$$

where (a) is by Fano's inequality, (b) is by adding and subtracting the same term, (c) is by applying Fano's inequality on the third term, and noticing that Lemma 1 together with the fact of the channel being discrete implies that,

$$\begin{aligned}
&\sum_{i=1}^{3} H(\mathcal{X}_i^n|\mathcal{W}_i) \\
&\geq H(\overline{\mathcal{X}}_1^n|\overline{\mathcal{W}}_1) + H(\overline{\mathcal{X}}_{12}^n|\overline{\mathcal{W}}_2) + H(X_{123}^n|W_{123}) \\
&\geq \max\{H(X_{123}^n|W_{123}), H(\overline{\mathcal{X}}_2^n|\overline{\mathcal{W}}_2)\} \qquad (26)
\end{aligned}$$

and (d) is again by the inequalities in (26). This completes the proof for the first rate combination. For the second rate combination, we have

$$\begin{aligned}
&6nR_1^* + 6nR_2^* + 3nR_3^* \\
&\geq \sum_{i=1}^{3} H(X_i^n|\overline{\mathcal{X}}_2^n\overline{\mathcal{W}}_2) + \sum_{i=1}^{3} H(\mathcal{X}_i^n) \\
&\overset{(a)}{\geq} \sum_{i=1}^{3} H(X_i^n|\overline{\mathcal{X}}_2^n\overline{\mathcal{W}}_2) + 3nR_1 + 6nR_2 + 3nR_3 \\
&\quad + H(\overline{\mathcal{X}}_2^n|\overline{\mathcal{W}}_2) - n\delta \\
&= \sum_{i=1}^{3} H(X_i^n\overline{\mathcal{X}}_2^n|\overline{\mathcal{W}}_2) - 3H(\overline{\mathcal{X}}_2^n|\overline{\mathcal{W}}_2) \\
&\quad + 3nR_1 + 6nR_2 + 3nR_3 + H(\overline{\mathcal{X}}_2^n|\overline{\mathcal{W}}_2) - n\delta \\
&\geq 6nR_1 + 6nR_2 + 3nR_3 - n\delta' \\
&\quad + \left[\sum_{i=1}^{3} H(X_i^n\overline{\mathcal{X}}_2^n|\mathcal{W}_i\overline{\mathcal{W}}_2) - 2H(\overline{\mathcal{X}}_2^n|\overline{\mathcal{W}}_2)\right] \\
&\overset{(b)}{\geq} 6nR_1 + 6nR_2 + 3nR_3 - n\delta', \qquad (27)
\end{aligned}$$

where (a) is because of (26), and in (b) we applied Lemma 1,

$$\sum_{i=1}^{3} H(X_i^n\overline{\mathcal{X}}_2^n|\mathcal{W}_i\overline{\mathcal{W}}_2) \geq H(\overline{\mathcal{X}}_1^n|\overline{\mathcal{W}}_1) + 2H(\overline{\mathcal{X}}_2^n|\overline{\mathcal{W}}_2), \quad (28)$$

and then omit the first term since the channel is discrete; the rest of the inequalities in (27) are by Fano's inequality. ■

This proof illustrates several main components of the proof for the general case. Firstly, the rate combination needs to be written as summations under appropriate proportions, secondly the K-way submodularity lemma needs to strategically used, and thirdly there are connections between different layers of messages and thus terms may be canceled among them. For the general K-user problem, the bounding becomes much more complicated, and we will rely on the optimal solution $B_{C^*}(\boldsymbol{A})$ to provide necessary structure and guidance.

B. Several Properties of $\phi_{i,j}$

We begin with a few properties on the exchange rate $\phi_{i,j}$.

Lemma 2: For any integers i, j, k such that $1 \leq i < j < k \leq K$, we have $\phi_{i,j}\phi_{j,k} = \phi_{i,k}$.

Lemma 3: For any integers i, j, k such that $1 \leq i < j < k \leq K$, we have $\phi_{k,j}\phi_{j,i} = \phi_{k,i}$.

Lemma 4: For any integers i, j such that $1 \leq i < j \leq K$, we have $\phi_{i,j}\phi_{j,i} = i/j < 1$.

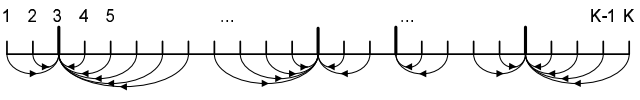

Fig. 4. An illustration of the optimal extremal solution structure. The longer and bolder marks give the set $\mathcal{E}$.

Lemma 5: For any integers i, j, k, we have $\phi_{i,k} \geq \phi_{i,j}\phi_{j,k}$, with equality only when the sequence (i, j, k) is monotonic.

Lemma 6: For any $k > j$, we have $(k+1)\binom{K-1}{k-1}\phi_{k+1,j} = k\binom{K-1}{k}\phi_{k,j}$.

Lemma 7: For any $i < j$, $\binom{K-1}{i-1}\binom{K-1}{j-1}^{-1} = \phi_{i,j}$.

The above lemmas (particularly Lemma 2-5) may be best understood as a currency exchange system where up-converting (or down-converting) many times results in the same final exchange rate as a single step conversion, but up-converting mixed with down-converting to the original currency results in a loss. The proofs of these lemmas are given in Appendix B.

C. Extremal Solutions and the Effective Rate Set

To prove the converse part of Theorem 2, we proceed in two steps: first we identify some special optimal solutions for the maximization problem (22) with certain desired properties, then show that $B_{\mathcal{C}^*}(\boldsymbol{A})$ is an upper bound to the quantity $B_{\mathcal{C}}(\boldsymbol{A})$. In this subsection we discuss the first step.

Definition 4: A non-negative setting of $r_{i,j}$ satisfying (15) is called extremal if the following conditions hold (i) For each $i = 1, 2, \ldots, K$, there exists a unique $j \in \mathcal{I}_K$ such that $r_{i,j} = R_i^*$ and $r_{i,k} = 0$ for $k \neq j$. (ii) If $r_{i,j} = R_i^* > 0$, then $r_{j,i} = R_j^*$. (iii) If $r_{i,j} = R_i^* > 0$, then for any k such that $\max(i, j) > k > \min(i, j)$, $r_{k,j} = R_k^*$.

Lemma 8: The solutions to the maximization problem (22) include one that is extremal.

The lemma is intuitively true since a linear optimization problem has an optimal solution at its corner point. The concept of extremal solution makes the definition of corner point in the problem context more precise. A proof is given in Appendix B.

Definition 5: In an optimal extremal solution, the effective rate set is defined as $\mathcal{E} \triangleq \{i \in \mathcal{I}_K : r_{j,i} > 0 \text{ for some } j\}$. The elements of $\mathcal{E}$ in an increasing order are denoted as $e_1, e_2, \ldots, e_{|\mathcal{E}|}$.

Lemma 8 implies there exists a specific structure of rate exchange in the optimal extremal solutions.

Lemma 9: For an optimal extremal solution:

- There exist a partition of the sequence $1, 2, \ldots, K$, labeled as $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{|\mathcal{E}|}$, each consisting a consecutive sequence of integers, and $e_i \in \mathcal{S}_i$.
- For $k \in \mathcal{S}_i$, we have $r_{k,e_i} = R_k^*$.

This structure is analogous to scalar quantization to some extent, as illustrated in Fig. 4.

D. Proof of the Converse Part of Theorem 2

Proof: For a fixed vector $\boldsymbol{A}$, let $\{\hat{r}_{i,j}\}$ be an optimal extremal solution for the maximization problem (22), and let $\mathcal{E}$

be its effective rate set and let $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{|\mathcal{E}|}$ be the partition sets; for convenience, denote the smallest element in the set $\mathcal{S}_i$ as l_i and the largest element as u_i. Assuming a sequence of length-n codes is given with diminishing error probability. Let $\mathcal{X}_i$ and $\overline{\mathcal{X}}_i$ be defined similarly as $\mathcal{W}_i$ and $\overline{\mathcal{W}}_i$. The proof consists of two layers of inductions. We start from the inner layer, and then put the pieces together in the outer layer.

Define the following quantity for $k = 1, 2, \ldots, |\mathcal{E}|$, for which lower bounds will be derived

$$L_k \triangleq \sum_{j=e_k}^{u_k-1} a_{k,j} \sum_{i=1}^{K} H(\mathcal{X}_i^n|\overline{\mathcal{X}}_{j+1}^n\overline{\mathcal{W}}_{j+1}) + a_{k,u_k} \sum_{i=1}^{K} H(\mathcal{X}_i^n|\overline{\mathcal{X}}_{u_k+1}^n\overline{\mathcal{W}}_{u_k+1}), \quad (29)$$

where

$$a_{k,j} \triangleq \frac{\phi_{j,e_k}}{\binom{K-1}{j-1}} - \frac{\phi_{j+1,e_k}}{\binom{K-1}{j}},\ j = e_k, \ldots, u_k - 1$$
$$a_{k,u_k} \triangleq \frac{\phi_{u_k,e_k}}{\binom{K-1}{u_k-1}} - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}}, \quad (30)$$

and for convenience we have defined $A_{e_{|\mathcal{E}|+1}} \triangleq 0$ and $\phi_{j,e_{|\mathcal{E}|+1}} \triangleq 0$. Note that all the coefficients in front of the entropy functions are non-negative: those in the first summation are straightforward to verify by using the definition of $\phi_{i,j}$, and for the last term we only need to observe that $A_{e_k}\phi_{u_k,e_k} \geq A_{e_{k+1}}\phi_{u_k,e_{k+1}}$ by the optimality of the extremal solution. For convenience let us also define

$$b_{k,j} \triangleq \frac{\phi_{j,e_k}}{\binom{K-1}{j-1}} - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}} = \sum_{i=j}^{u_k} a_{k,i},\ j = e_k, \ldots, u_k, \quad (31)$$

which are clearly non-negative quantities. We are interested in these quantities L_k's because they are directly related with the rate combination being considered, as we shall see shortly.

We start by writing the following

$$\begin{aligned}
&\sum_{i=1}^{K} H(\mathcal{X}_i^n|\overline{\mathcal{X}}_{j+1}^n\overline{\mathcal{W}}_{j+1}) \\
&= \sum_{i=1}^{K} H(\mathcal{X}_i^n|\overline{\mathcal{X}}_{j+1}^n\overline{\mathcal{W}}_{j+1}) + KH(\overline{\mathcal{X}}_{j+1}^n|\overline{\mathcal{W}}_{j+1}) \\
&\qquad - KH(\overline{\mathcal{X}}_{j+1}^n|\overline{\mathcal{W}}_{j+1}) \\
&= \sum_{i=1}^{K} H(\mathcal{X}_i^n, \overline{\mathcal{X}}_{j+1}^n|\overline{\mathcal{W}}_{j+1}) - KH(\overline{\mathcal{X}}_{j+1}^n|\overline{\mathcal{W}}_{j+1}) \\
&\overset{(a)}{\geq} \sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{j+1}) + \sum_{i=1}^{K} H(\mathcal{X}_i^n, \overline{\mathcal{X}}_{j+1}^n|\overline{\mathcal{W}}_{j+1}, \mathcal{W}_i) \\
&\qquad - KH(\overline{\mathcal{X}}_{j+1}^n|\overline{\mathcal{W}}_{j+1}) - n\delta \\
&\overset{(b)}{\geq} \sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{j+1}) + \sum_{i=1}^{j} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) \\
&\qquad - jH(\overline{\mathcal{X}}_{j+1}^n|\overline{\mathcal{W}}_{j+1}) - n\delta, \quad (32)
\end{aligned}$$

where (a) is by Fano's inequality, and (b) is by applying Lemma 1 on the second term. For notational simplicity, we shall ignore the small quantity δ in the sequel.

Slightly further expanding the first term in (32) and substituting it in L_k give us (33). More generally, we claim that for m such that $u_k - 1 \geq m \geq e_k - 1$, (34) holds, which we prove by induction. Clearly it holds for $m = u_k - 1$ since it is exactly (33) in this case. Suppose it holds for $m = m^*$, we shall prove it also holds for $m = m^* - 1$. Putting (32) into (34), we have (35) given on the next page. In order to simplify (35), first notice that $a_{k,m^*} + b_{k,m^*+1} = b_{k,m^*}$, and

$$\begin{aligned}
b_{k,m^*}\binom{K-1}{m^*-1} &= \phi_{m^*,e_k} - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}\binom{K-1}{m^*-1}}{A_{e_k}\binom{K-1}{u_k-1}} \\
&\overset{(a)}{=} \phi_{m^*,e_k} - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}\phi_{m^*,u_k}}{A_{e_k}} \\
&\overset{(b)}{=} \phi_{m^*,e_k} - \frac{A_{e_{k+1}}\phi_{m^*,e_{k+1}}}{A_{e_k}}, \quad (36)
\end{aligned}$$

where (a) is by Lemma 7 and (b) is by Lemma 2. It follows that

$$\begin{aligned}
&a_{k,m^*}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{m^*+1}) + b_{k,m^*+1}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{m^*+1}) \\
&= b_{k,m^*}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{m^*}) + nKb_{k,m^*}\binom{K-1}{m^*-1}R_{m^*}, \\
&= b_{k,m^*}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{m^*}) \\
&\qquad + nK\left(\phi_{m^*,e_k} - \frac{A_{e_{k+1}}\phi_{m^*,e_{k+1}}}{A_{e_k}}\right)R_{m^*}. \quad (37)
\end{aligned}$$

Furthermore, notice that

$$\begin{aligned}
&a_{k,m^*}\sum_{i=1}^{m^*} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) - a_{k,m^*}m^*H(\overline{\mathcal{X}}_{m^*+1}^n|\overline{\mathcal{W}}_{m^*+1}) \\
&\qquad + b_{k,m^*+1}\sum_{i=1}^{m^*+1} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) \\
&= b_{k,m^*}\sum_{i=1}^{m^*} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) \\
&\qquad + (b_{k,m^*+1} - a_{k,m^*}m^*)H(\overline{\mathcal{X}}_{m^*+1}^n|\overline{\mathcal{W}}_{m^*+1}) \\
&= b_{k,m^*}\sum_{i=1}^{m^*} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) \\
&\qquad - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}}H(\overline{\mathcal{X}}_{m^*+1}^n|\overline{\mathcal{W}}_{m^*+1}), \quad (38)
\end{aligned}$$

where the last step is due to

$$\begin{aligned}
&b_{k,m^*+1} - a_{k,m^*}m^* \\
&= (m^*+1)\frac{\phi_{m^*+1,e_k}}{\binom{K-1}{m^*}} - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}} - m^*\frac{\phi_{m^*,e_k}}{\binom{K-1}{m^*-1}} \\
&= -\frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}}, \quad (39)
\end{aligned}$$

where the last equality is by Lemma 6. Combining (35), (37) and (38), we have (40), proving that the claim (34) is indeed

true.

Letting $m = e_k - 1$, we can write (41) on this page. By breaking the second term as given in (42), where in the last step we apply Lemma 7 and Lemma 2, and noticing that for the third term

$$b_{k,e_k} \sum_{i=1}^{e_k} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i) \geq b_{k,e_k} \sum_{i=1}^{l_k} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i),$$

implied by the discrete nature of the channel, we can further write (43) on the next page.

This concludes the inner layer induction, and next we turn to the outer layer. First notice that the optimality of extremal solution and Lemma 9 together imply that

$$B_{C^*}(\boldsymbol{A}) = \sum_{i=1}^{|\mathcal{E}|} A_{e_i} \sum_{j \in \mathcal{S}_i} \phi_{j,e_i} R_j^*. \tag{44}$$

We first write (45) where the inequality can be justified as follows. Observe that in the second summation, for any $k > e_{|\mathcal{E}|}$, the random variables $X_{\mathcal{A}}$ with $|\mathcal{A}| = k$ appear only in the last $K - k + 1$ terms in the outer summation. Each inner summation has a total of $K\binom{K-1}{j-1}$ such terms, which implies such random variables are counted a total of $A_{e_{|\mathcal{E}|}} K \phi_{j,e_{|\mathcal{E}|}}$ times. Thus by the cardinality of the alphabets, the normalized entropy is upper bounded by R_k^*. Through a similar argument, it is not difficult to verify that for $k \leq e_{|\mathcal{E}|}$, all the terms are accounted for. Furthermore, notice that by the optimality of the extremal solution, for any $j \in \mathcal{S}_i$, we have $A_{e_i} \phi_{j,e_i} \geq A_{e_{|\mathcal{E}|}} \phi_{j,e_{|\mathcal{E}|}}$, and thus the first summation is non-negative.

We next apply (43) with $k = |\mathcal{E}|$ in (45), and write (46) on the next page, because we have $u_{|\mathcal{E}|} = K$, $A_{e_{|\mathcal{E}|+1}} = 0$ by definition, and

$$b_{|\mathcal{E}|,e_{|\mathcal{E}|}} = \frac{\phi_{e_{|\mathcal{E}|},e_{|\mathcal{E}|}}}{\binom{K-1}{e_{|\mathcal{E}|}-1}} = \frac{1}{\binom{K-1}{e_{|\mathcal{E}|}-1}}. \tag{47}$$

More generally, we claim for $k = 0, 1, \ldots, |\mathcal{E}| - 1$, the following inequality holds

$$\begin{aligned} B_{C^*}(\boldsymbol{A}) \geq & \sum_{i=1}^{k} \sum_{j \in \mathcal{S}_i} \left[A_{e_i} \phi_{j,e_i} - A_{e_{k+1}} \phi_{j,e_{k+1}} \right] R_j^* \\ & + \sum_{i=k+1}^{|\mathcal{E}|} A_{e_i} \sum_{j=l_i}^{u_i} \phi_{j,e_i} R_j + \frac{A_{e_{k+1}}}{nK\binom{K-1}{e_{k+1}-1}} \sum_{i=1}^{K} H(\mathcal{W}_i | \overline{\mathcal{W}}_{l_{k+1}}) \\ & + \frac{A_{e_{k+1}}}{nK\binom{K-1}{e_{k+1}-1}} \sum_{i=1}^{l_{k+1}} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i) \end{aligned} \tag{48}$$

We again take an induction approach to prove this claim. The claim is clearly true for $k = |\mathcal{E}| - 1$. Now suppose (48) is true for $k = k^*$, and we seek to show it is also true for $k = k^* - 1$. For notational simplicity, let us define

$$c_{j,k} = A_{e_i} \phi_{j,e_i} - A_{e_{k+1}} \phi_{j,e_{k+1}}. \tag{49}$$

We first prove the following inequality

$$\sum_{i=1}^{k^*} \sum_{j \in \mathcal{S}_i} c_{j,k^*} R_j^* \geq \sum_{i=1}^{k^*-1} \sum_{j \in \mathcal{S}_i} c_{j,k^*-1} R_j^* + \frac{A_{e_{k^*}}}{nK} L_{k^*}. \tag{50}$$

To do this, we need to count in the second term the number of appearance of random variables $X_{\mathcal{A}}$ for all $|\mathcal{A}| = m$, for all fixed m, such that $m \in \mathcal{I}_{u_{k^*}}$. This is similar to (45), but slightly more involved. For m such that $u_{k^*} \geq m > e_{k^*}$, it is easily seen that there are a total of $b_{k^*,m} K\binom{K-1}{m-1}$ such random variables in L_{k^*}, implying the following amount of R_m^* is accounted for

$$\frac{A_{e_{k^*}}}{K} b_{k^*,m} K \binom{K-1}{m-1} = A_{e_{k^*}} \phi_{m,e_{k^*}} - A_{e_{k^*+1}} \phi_{m,e_{k^*+1}}, \tag{51}$$

where we have used (36). This indeed is the difference between the left hand side of (50) and the first term on the right hand side, in terms of R_m^*. For the case $m \leq e_{k^*}$, the following amount of R_m^* is accounted for

$$\frac{A_{e_{k^*}}}{K} b_{k^*,e_{k^*}} K \binom{K-1}{m-1} = A_{e_{k^*}} \phi_{m,e^{k^*}} - A_{e_{k^*+1}} \phi_{m,e_{k^*+1}}, \tag{52}$$

where we have used the derivation in (42). This is again

$$\begin{aligned} L_k \geq & \sum_{j=e_k}^{u_k-1} a_{k,j} \sum_{i=1}^{K} H(\mathcal{X}_i^n | \overline{\mathcal{X}}_{j+1}^n \overline{\mathcal{W}}_{j+1}) + nK \left(\phi_{u_k,e_k} - \frac{A_{e_{k+1}} \phi_{u_k,e_{k+1}}}{A_{e_k}} \right) R_{u_k} \\ & + a_{k,u_k} \sum_{i=1}^{K} H(\mathcal{W}_i | \overline{\mathcal{W}}_{u_k}) + a_{k,u_k} \sum_{i=1}^{u_k} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i) - u_k a_{k,u_k} H(\overline{\mathcal{X}}_{u_k+1}^n | \overline{\mathcal{W}}_{u_k+1}). \end{aligned} \tag{33}$$

$$\begin{aligned} L_k \geq & \sum_{j=e_k}^{m} a_{k,j} \sum_{i=1}^{K} H(\mathcal{X}_i^n | \overline{\mathcal{X}}_{j+1}^n \overline{\mathcal{W}}_{j+1}) + nK \sum_{j=m+1}^{u_k} \left(\phi_{j,e_k} - \frac{A_{e_{k+1}} \phi_{j,e_{k+1}}}{A_{e_k}} \right) R_j + b_{k,m+1} \sum_{i=1}^{K} H(\mathcal{W}_i | \overline{\mathcal{W}}_{m+1}) \\ & + b_{k,m+1} \sum_{i=1}^{m+1} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i) - u_k a_{k,u_k} H(\overline{\mathcal{X}}_{u_k+1}^n | \overline{\mathcal{W}}_{u_k+1}) - \frac{A_{e_{k+1}} \phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}} \sum_{i=m+1}^{u_k-1} H(\overline{\mathcal{X}}_{i+1}^n | \overline{\mathcal{W}}_{i+1}). \end{aligned} \tag{34}$$

$$L_k \geq \sum_{j=e_k}^{m^*-1} a_{k,j} \sum_{i=1}^{K} H(\mathcal{X}_i^n|\overline{\mathcal{X}}_{j+1}^n\overline{\mathcal{W}}_{j+1}) + a_{k,m^*}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{m^*+1}) + a_{k,m^*}\sum_{i=1}^{m^*} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i)$$
$$- a_{k,m^*}m^* H(\overline{\mathcal{X}}_{m^*+1}^n|\overline{\mathcal{W}}_{m^*+1}) + nK\sum_{j=m^*+1}^{uk}\left(\phi_{j,e_k} - \frac{A_{e_{k+1}}\phi_{j,e_{k+1}}}{A_{e_k}}\right)R_j + b_{k,m^*+1}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{m^*+1})$$
$$+ b_{k,m^*+1}\sum_{i=1}^{m^*+1} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) - u_k a_{k,u_k} H(\overline{\mathcal{X}}_{u_k+1}^n|\overline{\mathcal{W}}_{u_k+1}) - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}}\sum_{i=m^*+1}^{u_k-1} H(\overline{\mathcal{X}}_{i+1}^n|\overline{\mathcal{W}}_{i+1}). \quad (35)$$

$$L_k \geq \sum_{j=e_k}^{m^*-1} a_{k,j}\sum_{i=1}^{K} H(\mathcal{X}_i^n|\overline{\mathcal{X}}_{j+1}^n\overline{\mathcal{W}}_{j+1}) + nK\sum_{j=m^*}^{u_k}\left(\phi_{j,e_k} - \frac{A_{e_{k+1}}\phi_{j,e_{k+1}}}{A_{e_k}}\right)R_j + b_{k,m^*}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{m^*})$$
$$+ b_{k,m^*}\sum_{i=1}^{m^*} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) - u_k a_{k,u_k} H(\overline{\mathcal{X}}_{u_k+1}^n|\overline{\mathcal{W}}_{u_k+1}) - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}}\sum_{i=m^*}^{u_k-1} H(\overline{\mathcal{X}}_{i+1}^n|\overline{\mathcal{W}}_{i+1}), \quad (40)$$

$$L_k \geq nK\sum_{j=e_k}^{u_k}\left(\phi_{j,e_k} - \frac{A_{e_{k+1}}\phi_{j,e_{k+1}}}{A_{e_k}}\right)R_j + b_{k,e_k}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{e_k}) + b_{k,e_k}\sum_{i=1}^{e_k} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i)$$
$$- u_k a_{k,u_k} H(\overline{\mathcal{X}}_{u_k+1}^n|\overline{\mathcal{W}}_{u_k+1}) - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}}\sum_{i=e_k}^{u_k-1} H(\overline{\mathcal{X}}_{i+1}^n|\overline{\mathcal{W}}_{i+1}). \quad (41)$$

$$b_{k,e_k}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{e_k}) = b_{k,e_k}nK\sum_{j=l_k}^{e_k-1}\binom{K-1}{j-1}R_j + b_{k,e_k}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_k})$$
$$= nK\sum_{j=l_k}^{e_k-1}\left(\frac{\binom{K-1}{j-1}}{\binom{K-1}{e_k-1}} - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}\binom{K-1}{j-1}}{A_{e_k}\binom{K-1}{u_k-1}}\right)R_j + b_{k,e_k}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_k})$$
$$= nK\sum_{j=l_k}^{e_k-1}\left(\phi_{j,e_k} - \frac{A_{e_{k+1}}\phi_{j,e_{k+1}}}{A_{e_k}}\right)R_j + b_{k,e_k}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_k}), \quad (42)$$

precisely the difference between the left hand side of (50) and the first term on the right hand side, in terms of R_m^*. Thus (50) is indeed true.

Now we proceed with the proof of (48) through induction by assuming it holds for $k = k^*$, and write (53) on the top of this page by applying (43). In order to simplify (53), similar terms need to be combined, for which we write (54), where in (a) we used Lemma 7, and (b) is because

$$\frac{A_{e_{k^*}}}{nK}b_{e_{k^*},e_{k^*}} + \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}}$$
$$= \frac{A_{e_{k^*}}}{nK\binom{K-1}{k^*-1}} - \frac{A_{e_{k^*+1}}\phi_{u_k,e_{k^*+1}}}{nK\binom{K-1}{u_{k^*}-1}} + \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}}$$
$$= \frac{A_{e_{k^*}}}{nK\binom{K-1}{k^*-1}} - \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{u_{k^*}-1}}\left(\phi_{u_k,e_{k^*+1}} - \frac{\binom{K-1}{u_{k^*}-1}}{\binom{K-1}{e_{k^*+1}-1}}\right)$$
$$= \frac{A_{e_{k^*}}}{nK\binom{K-1}{k^*-1}}, \quad (55)$$

$$L_k \geq nK\sum_{j=l_k}^{u_k}\left(\phi_{j,e_k} - \frac{A_{e_{k+1}}\phi_{j,e_{k+1}}}{A_{e_k}}\right)R_j + b_{k,e_k}\sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_k}) + b_{k,e_k}\sum_{i=1}^{l_k} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i)$$
$$- u_k a_{k,u_k} H(\overline{\mathcal{X}}_{u_k+1}^n|\overline{\mathcal{W}}_{u_k+1}) - \frac{A_{e_{k+1}}\phi_{u_k,e_{k+1}}}{A_{e_k}\binom{K-1}{u_k-1}}\sum_{i=e_k}^{u_k-1} H(\overline{\mathcal{X}}_{i+1}^n|\overline{\mathcal{W}}_{i+1}). \quad (43)$$

$$B_{C^*}(\boldsymbol{A}) \geq \sum_{i=1}^{|\mathcal{E}|-1} \sum_{j\in\mathcal{S}_i} \left[A_{e_i}\phi_{j,e_i} - A_{e_{|\mathcal{E}|}}\phi_{j,e_{|\mathcal{E}|}}\right] R_j^* + \frac{A_{e_{|\mathcal{E}|}}}{nK} \sum_{j=e_{|\mathcal{E}|}}^{K} \left(\frac{\phi_{j,e_{|\mathcal{E}|}}}{\binom{K-1}{j-1}} - \frac{\phi_{j+1,e_{|\mathcal{E}|}}}{\binom{K-1}{j}} \right) \sum_{i=1}^{K} H(\mathcal{X}_i^n | \overline{\mathcal{X}}_{j+1}^n \overline{\mathcal{W}}_{j+1})$$
$$= \sum_{i=1}^{|\mathcal{E}|-1} \sum_{j\in\mathcal{S}_i} \left[A_{e_i}\phi_{j,e_i} - A_{e_{|\mathcal{E}|}}\phi_{j,e_{|\mathcal{E}|}}\right] R_j^* + \frac{A_{e_{|\mathcal{E}|}}}{nK} L_{|\mathcal{E}|}, \tag{45}$$

$$B_{C^*}(\boldsymbol{A}) \geq \sum_{i=1}^{|\mathcal{E}|-1} \sum_{j\in\mathcal{S}_i} \left[A_{e_i}\phi_{j,e_i} - A_{e_{|\mathcal{E}|}}\phi_{j,e_{|\mathcal{E}|}}\right] R_j^* + A_{e_{|\mathcal{E}|}} \sum_{j=l_{|\mathcal{E}|}}^{u_{|\mathcal{E}|}} \left(\phi_{j,e_{|\mathcal{E}|}} - \frac{A_{e_{|\mathcal{E}|+1}}\phi_{j,e_{|\mathcal{E}|+1}}}{A_{e_{|\mathcal{E}|}}} \right) R_j$$
$$+ \frac{A_{e_{|\mathcal{E}|}}}{nK} b_{|\mathcal{E}|,e_{|\mathcal{E}|}} \left(\sum_{i=1}^{K} H(\mathcal{W}_i | \overline{\mathcal{W}}_{l_{|\mathcal{E}|}}) + \sum_{i=1}^{l_{|\mathcal{E}|}} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i) \right) - \frac{A_{e_{|\mathcal{E}|}}}{nK} u_{|\mathcal{E}|} a_{|\mathcal{E}|,u_{|\mathcal{E}|}} H(\overline{\mathcal{X}}_{u_{|\mathcal{E}|}+1}^n | \overline{\mathcal{W}}_{u_{|\mathcal{E}|}+1})$$
$$- \frac{A_{e_{|\mathcal{E}|}}}{nK} \frac{A_{e_{|\mathcal{E}|+1}}\phi_{u_{|\mathcal{E}|},e_{|\mathcal{E}|+1}}}{A_{e_{|\mathcal{E}|}}\binom{K-1}{u_{|\mathcal{E}|}-1}} \sum_{i=e_{|\mathcal{E}|}}^{u_{|\mathcal{E}|}-1} H(\overline{\mathcal{X}}_{i+1}^n | \overline{\mathcal{W}}_{i+1})$$
$$= \sum_{i=1}^{|\mathcal{E}|-1} \sum_{j\in\mathcal{S}_i} \left[A_{e_i}\phi_{j,e_i} - A_{e_{|\mathcal{E}|}}\phi_{j,e_{|\mathcal{E}|}}\right] R_j^* + A_{e_{|\mathcal{E}|}} \sum_{j=l_{|\mathcal{E}|}}^{K} \phi_{j,e_{|\mathcal{E}|}} R_j + \frac{A_{e_{|\mathcal{E}|}}}{nK\binom{K-1}{e_{|\mathcal{E}|}-1}} \sum_{i=1}^{K} H(\mathcal{W}_i | \overline{\mathcal{W}}_{l_{|\mathcal{E}|}})$$
$$+ \frac{A_{e_{|\mathcal{E}|}}}{nK\binom{K-1}{e_{|\mathcal{E}|}-1}} \sum_{i=1}^{l_{|\mathcal{E}|}} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i), \tag{46}$$

$$B_{C^*}(\boldsymbol{A}) \geq \sum_{i=1}^{k^*-1} \sum_{j\in\mathcal{S}_i} c_{j,k^*-1} R_j^* + A_{e_{k^*}} \sum_{j=l_{k^*}}^{u_{k^*}} \left(\phi_{j,e_{k^*}} - \frac{A_{e_{k^*+1}}\phi_{j,e_{k^*+1}}}{A_{e_{k^*}}} \right) R_j$$
$$+ \frac{A_{e_{k^*}}}{nK} b_{k^*,e_{k^*}} \left(\sum_{i=1}^{K} H(\mathcal{W}_i | \overline{\mathcal{W}}_{l_{k^*}}) + \sum_{i=1}^{l_{k^*}} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i) \right) - \frac{A_{e_{k^*}}}{nK} u_{k^*} a_{k^*,u_{k^*}} H(\overline{\mathcal{X}}_{u_{k^*}+1}^n | \overline{\mathcal{W}}_{u_{k^*}+1})$$
$$- \frac{A_{e_{k^*+1}}\phi_{u_{k^*},e_{k^*+1}}}{nK\binom{K-1}{u_{k^*}-1}} \sum_{i=e_{k^*}}^{u_{k^*}-1} H(\overline{\mathcal{X}}_{i+1}^n | \overline{\mathcal{W}}_{i+1}) + \sum_{i=k^*+1}^{|\mathcal{E}|} A_{e_i} \sum_{j=l_i}^{u_i} \phi_{j,e_i} R_j$$
$$+ \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} \sum_{i=1}^{K} H(\mathcal{W}_i | \overline{\mathcal{W}}_{l_{k^*+1}}) + \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} \sum_{i=1}^{l_{k^*+1}} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i). \tag{53}$$

where the last step is again by Lemma 7.

Next consider the summation (56) where we have split the last term and combined it with the other terms, and used (55); moreover, some terms $H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i)$ for $i = l_{k^*}+1, \ldots, e_{k^*}$ are ignored because they are non-negative by the discrete nature of the channel. Observe that for the last term in the right hand side of (56)

$$\frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} - \frac{A_{e_{k^*+1}}\phi_{u_{k^*},e_{k^*+1}}}{nK\binom{K-1}{u_{k^*}-1}}$$
$$= \frac{A_{e_{k^*+1}}\phi_{u_{k^*},e_{k^*+1}} - A_{e_{k^*+1}}\phi_{u_{k^*},e_{k^*+1}}}{nK\binom{K-1}{u_{k^*}-1}} = 0. \tag{57}$$

For the second term in the right hand side of (56), notice that $u_{k^*}+1 \in \mathcal{S}_{k^*+1}$, thus by the optimality of the extremal solution, we have

$$A_{e_{k^*+1}}\phi_{u_{k^*}+1,e_{k^*+1}} \geq A_{e_{k^*}}\phi_{u_{k^*}+1,e_{k^*}}, \tag{58}$$

and thus (59) follows, where (a) is by Lemma 4, and the final inequality is by (58). Now combining (53), (54), (56), (57) and (59) completes the induction proof of (48) for $k = k^* - 1$.

Writing (48) for $k = 0$, we have

$$B_{C^*}(\boldsymbol{A}) \geq \sum_{i=1}^{|\mathcal{E}|} A_{e_i} \sum_{j=l_i}^{u_i} \phi_{j,e_i} R_j$$
$$+ \frac{A_{e_1}}{nK\binom{K-1}{e_1-1}} \left(\sum_{i=1}^{K} H(\mathcal{W}_i | \overline{\mathcal{W}}_{l_1}) + \sum_{i=1}^{l_1} H(\overline{\mathcal{X}}_i^n | \overline{\mathcal{W}}_i) \right)$$
$$\geq \sum_{i=1}^{|\mathcal{E}|} A_{e_i} \sum_{j=l_i}^{u_i} \phi_{j,e_i} R_j, \tag{60}$$

$$
\begin{aligned}
&A_{e_{k^*}} \sum_{j=l_{k^*}}^{u_{k^*}} \left(\phi_{j,e_{k^*}} - \frac{A_{e_{k^*+1}}\phi_{j,e_{k^*+1}}}{A_{e_{k^*}}}\right) R_j + \frac{A_{e_{k^*}}}{nK} b_{k^*,e_{k^*}} \sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_{k^*}}) + \sum_{i=k^*+1}^{|\mathcal{E}|} A_{e_i} \sum_{j=l_i}^{u_i} \phi_{j,e_i} R_j \\
&\quad + \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} \sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_{k^*+1}}) \\
&= \sum_{j=l_{k^*}}^{u_{k^*}} \left(A_{e_{k^*}}\phi_{j,e_{k^*}} - A_{e_{k^*+1}}\phi_{j,e_{k^*+1}}\right) R_j + \frac{A_{e_{k^*}}}{nK} b_{k^*,e_{k^*}} \sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_{k^*}}) + \sum_{i=k^*+1}^{|\mathcal{E}|} A_{e_i} \sum_{j=l_i}^{u_i} \phi_{j,e_i} R_j \\
&\quad + \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} \sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_{k^*}}) + \sum_{j=l_{k^*}}^{u_{k^*}} \frac{A_{e_{k^*+1}}}{\binom{K-1}{e_{k^*+1}-1}} \binom{K-1}{j-1} R_j \\
&\overset{(a)}{=} \sum_{i=k^*}^{|\mathcal{E}|} A_{e_i} \sum_{j=l_i}^{u_i} \phi_{j,e_i} R_j + \left(\frac{A_{e_{k^*}}}{nK} b_{k^*,e_{k^*}} + \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}}\right) \sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_{k^*}}) \\
&\overset{(b)}{=} \sum_{i=k^*}^{|\mathcal{E}|} A_{e_i} \sum_{j=l_i}^{u_i} \phi_{j,e_i} R_j + \frac{A_{e_{k^*}}}{nK\binom{K-1}{e_{k^*}-1}} \sum_{i=1}^{K} H(\mathcal{W}_i|\overline{\mathcal{W}}_{l_{k^*}}),
\end{aligned} \tag{54}
$$

$$
\begin{aligned}
&\frac{A_{e_{k^*}}}{nK} b_{k^*,e_{k^*}} \sum_{i=1}^{l_{k^*}} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) - \frac{A_{e_{k^*}}}{nK} u_{k^*} a_{k^*,u_{k^*}} H(\overline{\mathcal{X}}_{u_{k^*}+1}^n|\overline{\mathcal{W}}_{u_{k^*}+1}) \\
&\quad - \frac{A_{e_{k^*+1}}\phi_{u_{k^*},e_{k^*+1}}}{nK\binom{K-1}{u_{k^*}-1}} \sum_{i=e_{k^*}}^{u_{k^*}-1} H(\overline{\mathcal{X}}_{i+1}^n|\overline{\mathcal{W}}_{i+1}) + \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} \sum_{i=1}^{l_{k^*+1}} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) \\
&\geq \frac{A_{e_{k^*}}}{nK\binom{K-1}{k^*-1}} \sum_{i=1}^{l_{k^*}} H(\overline{\mathcal{X}}_i^n|\overline{\mathcal{W}}_i) + \left(\frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} - \frac{A_{e_{k^*}}}{nK} u_{k^*} a_{k^*,u_{k^*}}\right) H(\overline{\mathcal{X}}_{u_{k^*}+1}^n|\overline{\mathcal{W}}_{u_{k^*}+1}) \\
&\quad + \left(\frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} - \frac{A_{e_{k^*+1}}\phi_{u_{k^*},e_{k^*+1}}}{nK\binom{K-1}{u_{k^*}-1}}\right) \sum_{i=e_{k^*}}^{u_{k^*}-1} H(\overline{\mathcal{X}}_{i+1}^n|\overline{\mathcal{W}}_{i+1}),
\end{aligned} \tag{56}
$$

$$
\begin{aligned}
\frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} - \frac{A_{e_{k^*}}}{nK} u_{k^*} a_{k^*,u_{k^*}} &= \frac{A_{e_{k^*+1}}}{nK\binom{K-1}{e_{k^*+1}-1}} - \frac{A_{e_{k^*}} u_{k^*}}{nK} \left(\frac{\phi_{u_{k^*},e_{k^*}}}{\binom{K-1}{u_{k^*}-1}} - \frac{A_{e_{k^*+1}}\phi_{u_{k^*},e_{k^*+1}}}{A_{e_{k^*}}\binom{K-1}{u_{k^*}-1}}\right) \\
&= \frac{(u_{k^*}+1)A_{e_{k^*+1}}\phi_{u_{k^*},e_{k^*+1}} - A_{e_{k^*}} u_{k^*}\phi_{u_{k^*},e_{k^*}}}{nK\binom{K-1}{u_{k^*}-1}} \\
&\overset{(a)}{=} \frac{A_{e_{k^*+1}}\phi_{u_{k^*}+1,e_{k^*+1}} - A_{e_{k^*}} u_{k^*}\phi_{u_{k^*},e_{k^*}}\phi_{u_{k^*}+1,u_{k^*}}}{nK\binom{K-1}{u_{k^*}-1}\phi_{u_{k^*}+1,u_{k^*}}} \geq 0,
\end{aligned} \tag{59}
$$

where the second inequality is because the first term in the parenthesis degenerates to zero, and the second is non-negative. Notice that for any $j \in \mathcal{S}_i$, by the optimality of the given extremal solution, $A_j \leq A_{e_i}\phi_{j,e_i}$, thus by the non-negativeness of rate R_i's, we arrive at

$$
B_{C^*}(\boldsymbol{A}) \geq \sum_{i=1}^{|\mathcal{E}|} A_{e_i} \sum_{j=l_i}^{u_i} \phi_{j,e_i} R_j \geq \sum_{i=1}^{K} A_i R_i. \tag{61}
$$

This completes the proof. ∎

For the case that $2^{R_i^*}$'s are not integers, we can instead consider a sequence of channels with memory, for which the alphabet sizes are $2^{nR_i^*}$, however, for each n channel use, the channel erases $(n-1)$ of them. This channel is not a memoryless channel anymore, however, our definition is sufficiently general to include such a case, and the converse proof can be used without any change.

VI. Conclusion

We consider the latent capacity region of the symmetric broadcast problem, which gives the maximum implication region for a specific achievable rate vector. A complete characterization is provided, for which the converse proof relies on a deterministic channel model, and deriving upper bounds for any bounding plane of the rate region. The forward

proof reveals an inherent connection between broadcast with common messages and erasure correction codes.

We believe the latent capacity region (or latent rate region) is a general concept, and can be applied to other problem. In [1], the multiple access channel is also considered for two and three-user case. It is conceivable that the technique used in this work can be used to generalize their results for the multiple access channel. Another interesting case may be the interference channel, where the well-known Han-Kobayashi region [9] is indeed the projection of a rate region for the coding problem with common messages. A careful analysis of the latent capacity region for the general interference channel may yield further insight into the problem.

ACKNOWLEDGMENT

The author wishes to thank the anonymous reviewers for their comments which help improve the presentation of this paper.

APPENDIX A
SUBMODULARITY PROPERTY OF CONDITIONAL ENTROPY

***Lemma** 10:* Let $U_1, U_2, \ldots, U_N$ be a set of mutually independent random variables, and let $V_1, V_2, \ldots, V_N$ be N random variables jointly distributed with them. Let $\mathcal{G}$ be a subset of $\mathcal{I}_N$, i.e., $\mathcal{G} \subseteq \mathcal{I}_N$. The conditional entropy function $H_{V|U}(\mathcal{G}) \triangleq H(V_i, i \in \mathcal{G}|U_i, i \in \mathcal{G})$ is a submodular function, i.e., for any $\mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{I}_N$,

$$H_{V|U}(\mathcal{G}_1) + H_{V|U}(\mathcal{G}_2) \geq H_{V|U}(\mathcal{G}_1 \cup \mathcal{G}_2) + H_{V|U}(\mathcal{G}_1 \cap \mathcal{G}_2).$$

Proof: Notice that

$$\begin{aligned} &H_{V|U}(\mathcal{G}_1) + H_{V|U}(\mathcal{G}_2) \\ &= H(V_i, U_i, i \in \mathcal{G}_1) + H(V_i, U_i, i \in \mathcal{G}_2) \\ &\quad - H(U_i, i \in \mathcal{G}_1) - H(U_i, i \in \mathcal{G}_2), \end{aligned} \tag{62}$$

and

$$\begin{aligned} &H_{V|U}(\mathcal{G}_1 \cup \mathcal{G}_2) + H_{V|U}(\mathcal{G}_1 \cap \mathcal{G}_2) \\ &= H(V_i, U_i, i \in \mathcal{G}_1 \cup \mathcal{G}_2) + H(V_i, U_i, i \in \mathcal{G}_1 \cap \mathcal{G}_2) \\ &\quad - H(U_i, i \in \mathcal{G}_1 \cup \mathcal{G}_2) - H(U_i, i \in \mathcal{G}_1 \cap \mathcal{G}_2). \end{aligned} \tag{63}$$

The mutual independence among U_i's gives

$$\begin{aligned} &H(U_i, i \in \mathcal{G}_1) + H(U_i, i \in \mathcal{G}_2) \\ &= H(U_i, i \in \mathcal{G}_1 \cup \mathcal{G}_2) + H(U_i, i \in \mathcal{G}_1 \cap \mathcal{G}_2). \end{aligned} \tag{64}$$

The submodularity of unconditioned entropy function of random variables is well-known [8], which gives

$$\begin{aligned} &H(V_i, U_i, i \in \mathcal{G}_1) + H(V_i, U_i, i \in \mathcal{G}_2) \\ &\geq H(V_i, U_i, i \in \mathcal{G}_1 \cup \mathcal{G}_2) + H(V_i, U_i, i \in \mathcal{G}_1 \cap \mathcal{G}_2) \end{aligned} \tag{65}$$

and the proof is thus complete. ■

APPENDIX B
PROOF OF THE LEMMAS

Proof of Lemma 2:

$$\begin{aligned} \phi_{i,j}\phi_{j,k} &= \binom{K-i}{j-i}^{-1}\binom{j-1}{j-i}\binom{K-j}{k-j}^{-1}\binom{k-1}{k-j} \\ &= \frac{(j-i)!(K-j)!(j-1)!}{(K-i)!(j-i)!(i-1)!}\frac{(k-j)!(K-k)!(k-1)!}{(K-j)!(k-j)!(j-1)!} \\ &= \frac{(K-k)!(k-1)!}{(K-i)!(i-1)!}\frac{(k-i)!}{(k-i)!} = \phi_{i,k}. \end{aligned} \tag{66}$$

■

Proof of Lemma 3:

$$\begin{aligned} \phi_{k,j}\phi_{j,i} &= \binom{k}{k-j}^{-1}\binom{K-j}{k-j}\binom{j}{j-i}^{-1}\binom{K-i}{j-i} \\ &= \frac{(k-j)!j!(K-j)!}{k!(k-j)!(K-k)!}\frac{(j-i)!i!(K-i)!}{j!(j-i)!(K-j)!} \\ &= \frac{i!(K-i)!}{k!(K-k)!}\frac{(k-i)!}{(k-i)!} = \phi_{k,i}. \end{aligned} \tag{67}$$

■

Proof of Lemma 4: By the definition of $\phi_{i,j}$, it is easy to verify that

$$\phi_{i,j}\phi_{j,i} = \frac{i}{j} < 1. \tag{68}$$

■

Proof of Lemma 5: The case $i = k$ is exactly Lemma 4, thus we only need to consider the case $i \neq k$; we may also assume $j \neq i$ and $j \neq k$ since these cases are trivial. The order of i, j, k can be arbitrary, but since the proof only relies on Lemma 2, 3 and 4, we may assume without loss of generality $i < j$. Thus we have the only three cases. (1) $k < i < j$: by Lemma 3 and 4, we have $\phi_{i,k} > \phi_{i,k}\phi_{i,j}\phi_{j,i} = \phi_{i,j}\phi_{j,k}$. (2) $i < k < j$: by Lemma 2 and 4, we have $\phi_{i,k} > \phi_{i,k}\phi_{k,j}\phi_{j,k} = \phi_{i,j}\phi_{j,k}$. (3) $i < j < k$: the equality is implied by Lemma 2.

■

Proof of Lemma 6:

$$\begin{aligned} &(k+1)\binom{K-1}{k-1}\phi_{k+1,j} \\ &= \frac{(k+1)(K-1)!(K-j)!(k+1-j)!j!}{(k-1)!(K-k)!(k+1-j)!(K-k-1)!(k+1)!} \\ &= \frac{(K-1)!(K-j)!j!}{(k-1)!(K-k)!(K-k-1)!k!}. \end{aligned} \tag{69}$$

Similarly, we have

$$\begin{aligned} k\binom{K-1}{k}\phi_{k,j} &= \frac{k(K-1)!(K-j)!(k-j)!j!}{k!(K-k-1)!(k-j)!(K-k)!k!} \\ &= \frac{(K-1)!(K-j)!j!}{(k-1)!(K-k)!(K-k-1)!k!}, \end{aligned} \tag{70}$$

proving the lemma. ■

Proof of Lemma 7: We only need to write the following

$$\begin{aligned}\binom{K-1}{i-1}\binom{K-1}{j-1}^{-1} &= \frac{(K-1)!(j-1)!(K-j)!}{(i-1)!(K-i)!(K-1)!} \\ &= \frac{(j-1)!}{(i-1)!(j-i)!}\frac{(K-j)!(j-i)!}{(K-i)!} \\ &= \binom{j-1}{j-i}\binom{K-i}{j-i}^{-1} = \phi_{i,j}. \quad (71)\end{aligned}$$

■

Proof of Lemma 8: Suppose an arbitrary optimal solution of the maximization problem (22) is given, we shall next transform it into an extremal one which is also optimal.

For condition (i), we may assume $R_i^* > 0$ because otherwise the statement is trivial. Observe that for any optimal solution, the second inequality in (16) must hold with equality, because otherwise the quantity being maximized can strictly increase. First suppose for certain i, there exist distinct j_1, j_2 such that $r_{i,j_1} > 0$ and $r_{i,j_2} > 0$, then we must have

$$A_{j_1}\phi_{i,j_1}r_{i,j_1} = A_{j_2}\phi_{i,j_2}r_{i,j_2}, \quad (72)$$

because otherwise, e.g., if $<$ held, then letting $r'_{i,j_2} = r_{i,j_1} + r_{i,j_2}$ and $r'_{i,j_1} = 0$ strictly increases the quantity being maximized in (22). However, if (72) is true, the new solution given above does not decrease the quantity being maximized, thus a new solution can be found such that there exist no such two distinct j_1, j_2. Given this is true, it is clear that for each i, letting the unique j for which $r_{i,j} > 0$ be R_i^* is an optimal choice. Thus condition (i) is satisfied by some optimal solution, and from here on, we shall only consider such solutions.

For condition (ii), we may assume $R_j^* > 0$ since otherwise the statement is trivial. Suppose condition (ii) is not true, i.e., for some $k \neq j$, $r_{j,k} = R_j^*$, then the optimality of the solution implies

$$A_j \leq A_k\phi_{j,k}, \quad (73)$$

Now we claim that the new solution with $r'_{i,k} = R_i^*$ and $r'_{i,j} = 0$ (with other $r_{i,j}$ values unchanged) can not decrease the quantity being maximized. To see this, we only need to observe that

$$A_k\phi_{i,k}R_i^* \geq A_k\phi_{i,j}\phi_{j,k}R_i^* \geq A_j\phi_{i,j}R_i^*, \quad (74)$$

which is by Lemma 5 and (73). Thus the conditions (i) and (ii) are indeed satisfied by some optimal solution, and from here on we shall only consider such solutions.

For condition (iii), we only discuss the case $i < j$, because the other case $i > j$ is similar. The fact that $r_{i,j} > 0$ implies $A_iR_i^* \leq A_j\phi_{i,j}R_i^*$. We may assume $R_k^* > 0$ because otherwise the statement is trivial. Take an arbitrary k, such that $i < k < j$, we may have $r_{k,j'} = R_k^*$ for some j', and the value of j' may be $j' < i$, $i < j' < k$ or $j' \leq k$; note that we can assume $j' \neq i$ since condition (i) afore-proved. It is easy to see that we must have $A_j > 0$ and $A_{j'} > 0$. The fact that $r_{i,j} > 0$ and $r_{k,j'} > 0$ imply that

$$A_j\phi_{i,j} \geq A_{j'}\phi_{i,j'}, \quad \text{and} \quad A_{j'}\phi_{k,j'} \geq A_j\phi_{k,j}. \quad (75)$$

The three cases are now discussed individually next. Case (1) $j' < i$, from (75) and Lemma 2 and 3, we have

$$A_j\phi_{i,k}\phi_{k,j} \geq A_{j'}\phi_{i,j'}, \quad \text{and} \quad A_{j'}\phi_{k,i}\phi_{i,j'} \geq A_j\phi_{k,j}, \quad (76)$$

which lead to $\phi_{i,k}\phi_{k,i} \geq 1$, contradicting Lemma 4, thus this is an impossible case. Case (2) $i < j' < k$, from (75) and Lemma 2, we have

$$A_j\phi_{i,j'}\phi_{j',k}\phi_{k,j} \geq A_{j'}\phi_{i,j'}, \quad \text{and} \quad A_{j'}\phi_{k,j'} \geq A_j\phi_{k,j}, \quad (77)$$

which lead to $\phi_{j',k}\phi_{k,j'} \geq 1$, thus this is another impossible case. Case (3) $j' \geq k$, from (75) and Lemma 2, we have that for this case

$$A_j\phi_{i,k}\phi_{k,j} \geq A_{j'}\phi_{i,k}\phi_{k,j'} \quad (78)$$

thus the new solution that $r'_{k,j} = r_{k,j'}$ and $r'_{k,j'} = 0$ does not decrease the quantity being optimized. Thus the conditions (i), (ii) and (iii) are indeed satisfied simultaneously by some optimal solution. The lemma is proved. ■

REFERENCES

[1] L. Grokop and D.N.C. Tse, "Fundamental constraints on multicast capacity regions," preprint, arXiv:0809.2835v1.

[2] C. Tian and S. Diggavi, "On multistage successive refinement for Wyner-Ziv source coding with degraded side information," *IEEE Trans. Information Theory*, vol. 53, no. 8, pp. 2946–2960, Aug. 2007.

[3] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Information Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.

[4] C. Tian and S. Diggavi, "Side-information scalable source coding," *IEEE Trans. Information Theory*, vol. 54, no. 12, pp. 5591–5608, Dec. 2008.

[5] H. J. A. Harvey, R. Kleinberg, and A. R. Lehman, "On the capacity of information networks," *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2345–2364, Jun. 2006.

[6] R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung, "Network information flow," *IEEE Trans. on Information Theory*, vol. 46, pp. 1004-1016, Jul. 2000.

[7] G. M. Ziegler, *Lectures on Polytopes, volume 152 of Graduate Texts in Mathematics*, Springer-Verlag, 1995.

[8] M. S. Fujishige, *Submodular functions and optimization, annals of discrete mathematics 47*, Elsevier Science Publishing Company, 1991.

[9] T. S. Han, and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Information Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.